(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 237 362**
A1

(12) # EUROPEAN PATENT APPLICATION

(21) Application number: 87302196.8

(22) Date of filing: 13.03.87

(51) Int. Cl.³: C 12 Q 1/68
C 12 Q 1/70

(30) Priority: 13.03.86 US 839331
22.08.86 US 899344

(43) Date of publication of application:
16.09.87 Bulletin 87/38

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(71) Applicant: CETUS CORPORATION
1400 Fifty-Third Street
Emeryville California 94608(US)

(72) Inventor: Erlich, Henry Anthony
3936 Rhoda Avenue
Oakland California 94602(US)

(72) Inventor: Saiki, Randall Keichi
320-39th Street
Richmond California 94805(US)

(72) Inventor: Horn, Glenn Thomas
7 Commodore Drive
Emeryville California 94608(US)

(72) Inventor: Mullis, Kary Banks
6767 Neptune Place No. 4
La Jolla California 92037(US)

(74) Representative: Bizley, Richard Edward et al,
BOULT, WADE & TENNANT 27 Furnival Street
London EC4A 1PQ(GB)

(54) **Process for detecting specific nucleotide variations and genetic polymorphisms present in nucleic acids and kits therefor.**

(57) Single or multiple nucleotide variations in nucleic acid sequence can be detected in nucleic acids by a process whereby the sample suspected of containing the relevant nucleic acid is repeatedly treated with primers, nucleotide triphosphates, and an agent for polymerization of the triphosphates and then denatured, in a process which amplifies the sequence containing the nucleotide variation if it is present. In one embodiment, the sample is spotted on a membrane and treated with a labeled sequence-specific oligonucleotide probe. Hybridization of the probe to the sample is detected by the label on the probe.

EP 0 237 362 A1

0237362

PROCESS FOR DETECTING SPECIFIC NUCLEOTIDE VARIATIONS AND GENETIC POLYMORPHISMS PRESENT IN NUCLEIC ACIDS AND KITS THEREFOR

This invention relates to a process for detecting nucleotide variations, mutations and polymorphisms by amplifying nucleic acid sequences suspected of containing such mutations or polymorphisms and detecting them with sequence-specific oligonucleotides in a dot blot format.

In recent years, the molecular basis of a number of human genetic diseases has been elucidated by the application of recombinant DNA technology. In particular, the detection of specific polymorphic restriction sites in human genomic DNA associated with genetic disease, such as sickle-cell anemia, has provided clinically valuable information for prenatal diagnosis. In these studies, the presence or absence of a specific site is revealed by restriction fragment length polymorphism (RFLP) analysis, a method in which variation in the size of a specific genomic restriction fragment is detected by Southern blotting and hybridization of the immobilized genomic DNA with a labeled probe. RFLP analysis has proved useful in the direct detection of polymorphic sites that contain the mutation conferring the disease phenotype (e.g., MstII and sickle-cell anemia) as well as in linkage studies where a particular allelic restriction site is linked to a disease locus within a family but not necessarily in the general population. See, for example, Kan and Dozy, PNAS (USA), 75, 5631 (1978), and Rubin and Kan, Lancet, 1985-I, 75 (1985). See also Geever et al., PNAS (USA), 78, 5081 (1981) and Wilson et al., PNAS (USA), 79, 3628 (1982).

In a second method, called "oligomer restriction", a synthetic end-labeled oligonucleotide probe is annealed in solution to the target genomic DNA sequence and a restriction enzyme is added to cleave any hybrids formed. This method, the specificity of which depends on the ability of a base pair mismatch within the restriction site to abolish or inhibit cleavage, is described more fully by Saiki et al., Biotechnology, 3, 1008-1012 (1985). In addition, the

sensitivity of this technique may be enhanced by utilizing a polymerase chain reaction procedure wherein the sample is first subjected to treatment with specific primers, polymerase and nucleotides to amplify the signal for subsequent detection. This is described more fully by Saiki et al., Science, 230, 1350-1353 (1985).

A third method for detecting allelic variations which is independent of restriction site polymorphism utilizes sequence-specific synthetic oligonucleotide probes. See Conner et al., PNAS (USA), 80, 78 (1983). This latter technique has been applied to the prenatal diagnosis of α1-antitrypsin deficiency (Kidd et al., Nature, 304, 230 (1983)) and β-thalassemia (Pirastu et al., N. Engl. J. Med., 309, 284 (1983)). In addition, the technique has been applied to study the polymorphism of HLA-DRβ using Southern blotting (Angelini et al., PNAS (USA), 83, 4489-4493 (1986)).

The basis for this procedure is that under appropriate hybridization conditions a short oligonucleotide probe of at least 19 bases (19-mer) will anneal only to those sequences to which it is perfectly matched, a single base pair mismatch being sufficiently destabilizing to prevent hybridization. The distinction between the allelic variants is based on the thermal stability of the duplex formed between the genomic DNA and the oligonucleotide (19-mer) probe.

In addition, methods for detecting base pair mismatches in double-stranded RNA and RNA:DNA heteroduplexes have been described using pancreatic ribonuclease (RNase A) to cleave the heteroduplexes. Winter et al., PNAS (USA), 82:7575-7579 (1985) and Myers et al., Science, 230:1242-1246 (1985). The principal deficiency of this method is its inability to recognize all types of base pair mismatches.

Both the RFLP and oligonucleotide probe stability methods are relatively complex procedures, requiring restriction enzyme digestion, gel-fractionation of the genomic DNA, denaturation of the DNA, immobilization of the DNA either by transfer to a filter membrane or dessication of the gel itself, and hybridization of a labeled probe to the electrophoretically resolved array of immobilized genomic

restriction fragments. These steps are necessary, for the oligonucleotide probe stability method, due to the complexity of human genomic DNA. Restriction and electrophoresis are necessary to separate the target sequence ("signal") from the rest of the genome ("noise"), and hybridization in the gel (instead of filter transfer) is necessary to retain as much target sequence as possible. Even then, detection of a signal in a 10 µg sample using a high specific activity kinased probe requires an autoradiographic exposure of four days.

In addition, the approach of Conner et al. requires at least a 19-mer probe for reasons of specificity (a shorter probe would hybridize to more genomic fragments), as well as possibly sensitivity. Shorter probes (e.g., 16-mers), however, would show more sequence-specific discrimination because a single mismatch would be more destabilizing.

There is a need in the art for a simplified method with improved sensitivity and specificity for directly detecting single-base differences in nucleic acids such as genomic DNA.

Accordingly, the present invention provides a method for detecting single or multiple nucleotide variations in nucleic acid sequence from any source, for use in detecting any type of disease or condition. The method herein directly detects the sequence variation, eliminating the need for restriction digestion, electrophoresis, and gel manipulations otherwise required. In addition, the method herein provides for improved specificity and sensitivity of the probe; an interpretable signal can be obtained with a 0.04 µg sample in six hours. Thirdly, if the amount of sample spotted on a membrane is increased to 0.1-0.5 µg, non-isotopically labeled oligonucleotides may be utilized rather than the radioactive probes used in previous methods. Finally, the process herein is applicable to use of sequence-specific oligonucleotides less than 19-mers in size, thus allowing use of more discriminatory sequence-specific oligonucleotides.

Specifically, the present invention provides for a process for detecting the presence or absence of at least one nucleotide variation in sequence in one or more nucleic acids contained in a sample, which process comprises:

(a) treating the sample, together or sequentially, with four different nucleotide triphosphates, an agent for polymerization of the nucleotide triphosphates, and one oligonucleotide primer for each strand of each nucleic acid suspected of containing said variation under hybridizing conditions, such that for each nucleic acid strand containing each different variation to be detected, an extension product of each primer is synthesized which is complementary to each nucleic acid strand, wherein said primer or primers are selected so as to be substantially complementary to each nucleic acid strand containing each different variation, such that the extension product synthesized from one primer, when it is separated from its complement, can serve as a template for synthesis of the extension product of the other primer;

(b) treating the sample under denaturing conditions to separate the primer extension products from their templates if the variation(s) to be detected are present;

(c) treating the sample, together or sequentially, with said four nucleotide triphosphates, an agent for polymerization of the nucleotide triphosphates, and oligonucleotide primers such that a primer extension product is synthesized using each of the single strands produced in step (b) as a template, wherein steps (b) and (c) are repeated a sufficient number of times to result in detectable amplification of the nucleic acid containing the sequence variation(s), if present;

(d) affixing the product of step (c) to a membrane;

(e) treating the membrane under hybridization conditions with a labeled sequence-specific oligonucleotide probe capable of hybridizing with the amplified nucleic acid sequence only if a sequence of the probe is complementary to a region of the amplified sequence; and

(f) detecting whether the probe has hybridized to an amplified sequence in the nucleic acid sample.

Steps (b) and (c) are preferably repeated at least five times, and more preferably 15-30 times if the sample contains human genomic DNA. If the sample comprises cells, preferably they are heated before step (a) to expose the nucleic acids therein to the reagents. This step avoids purification of the nucleic acids prior to reagent addition.

In a variation of this process, the primer(s) and/or nucleotide triphosphates are labeled so that the resulting amplified sequence is labeled. The labeled primer(s) and/or nucleotide triphosphate(s) can be present in the reaction mixture initially or added during a later cycle. The sequence-specific oligonucleotide (unlabeled) is affixed to a membrane and treated under hybridization conditions with the labeled amplification product so that hybridization will occur only if the membrane-bound sequence is present in the amplification product.

In another embodiment, the invention herein relates to a kit for detecting the presence or absence of at least one nucleotide variation in sequence in one or more nucleic acids contained in a sample, which kit comprises, in packaged form, a multicontainer unit having:

(a) one container for each oligonucleotide primer for each nucleic acid strand containing each different variation being detected, which primer(s) are substantially complementary to each strand containing each different variation, such that an extension product synthesized from one primer, when it is separated from its complement, can serve as a template for the synthesis of the extension product of the other primer so as to produce one or more amplified nucleic acid sequences if the sequence variation(s) are present;

(b) a container for an agent for polymerization;

(c) a container for each of four different nucleotide triphosphates;

(d) one container for each labeled sequence-specific oligonucleotide capable of hybridizing with each possible sequence variation in the amplified nucleic acid sequence; and

(e) container(s) for reagents which detect hybridization of the probes to the amplified sequence.

Regarding genetic diseases, while RFLP requires a polymorphic restriction site to be associated with the disease, sequence-specific oligonucleotides directly detect the genetic lesion and are generally more useful for the analysis of such genetic diseases as hemoglobin C disease, α1-antitrypsin and β-thalassemia which result from single or multiple base mutations. In addition, the oligonucleotides can be used to distinguish between genetic variants which represent different alleles (e.g., HLA typing), indicating the feasibility of a sequence-specific oligonucleotide-based HLA typing kit.

The term "nucleotide variation in sequence" refers to any single or multiple nucleotide substitutions, deletions or insertions. These nucleotide variations may be mutant or polymorphic allele variations. Therefore, the process herein can detect single nucleotide changes in nucleic acids such as occur in β-globin genetic diseases caused by single-base mutations, additions and deletions (some β-thalassemias, sickle cell anemia, hemoglobin C disease, etc.), as well as multiple-base variations such as are involved with α-thalassemia or some β-thalassemias. The process can also detect polymorphisms, which are not necessarily associated with a disease, but are merely a condition in which two or more different nucleotide sequences (whether having substituted, deleted or inserted nucleotide base pairs) can exist at a particular site in the nucleic acid in the population, as with HLA regions of the human genome and random polymorphisms such as in mitochondrial DNA. The polymorphic sequence-specific oligonucleotide probes described in detail hereinafter may be used as genetic markers linked to a disease such as insulin-dependent diabetes or in forensic applications. If the nucleic acid is double-stranded, the nucleotide variation in sequence becomes a base pair variation in sequence.

The term "oligonucleotide" as used herein is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than three. Its exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. The oligonucleotide may be derived synthetically or by cloning.

The term "primer" as used herein refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of four different nucleotide triphosphates and an agent for polymerization such as DNA polymerase in an appropriate buffer ("buffer" includes pH, ionic strength, cofactors, etc.) and at a suitable temperature.

The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. The exact lengths of the primers will depend on many factors, including temperature and source of primer and use of the method. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25 nucleotides, although it may contain more or fewer nucleotides. Short primer molecules generally require lower temperatures to form sufficiently stable hybrid complexes with the template.

The primers herein are selected to be "substantially" complementary to the different strands of each specific sequence to be amplified. This means that the primers must be sufficiently complementary to hybridize with their respective strands. Therefore,

the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Typically, the primers have exact complementarity to obtain the best detection results.

The term "sequence-specific oligonucleotides" refers to oligonucleotides which will hybridize to specific sequences, whether or not contained on alleles which sequences span the nucleotide variation being detected and are specific for the sequence variation being detected. Depending on the sequences being analyzed, one or more sequence-specific oligonucleotides may be employed for each sequence, as described further hereinbelow.

As used herein, the term "thermostable enzyme" refers to an enzyme which is stable to heat and is heat resistant and catalyzes (facilitates) combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and will proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be thermostable enzymes, however, which initiate synthesis at the 5' end and proceed in the other direction, using the same process as described above. A purified thermostable enzyme is described more fully in Example VIII hereinbelow.

The present invention is directed to a process for amplifying any one or more specific nucleic acid sequences (as defined herein to contain one or more nucleotide variations) suspected of being in one or more nucleic acids.

In general, the present process involves a chain reaction for producing, in exponential quantities relative to the number of reaction steps involved, at least one specific nucleic acid sequence given (a) that the ends of the required sequence are known in sufficient detail that oligonucleotides can be synthesized which will hybridize to them, and (b) that a small amount of the sequence is

available to initiate the chain reaction. The product of the chain reaction will be a discrete nucleic acid duplex with termini corresponding to the ends of the specific primers employed.

Any nucleic acid, in purified or nonpurified form, can be utilized as the starting nucleic acid or acids, provided it is suspected of containing the sequence being detected. Thus, the process may employ, for example, DNA or RNA, including messenger RNA, which DNA or RNA may be single stranded or double stranded. In addition, a DNA-RNA hybrid which contains one strand of each may be utilized. A mixture of any of these nucleic acids may also be employed, or the nucleic acids produced from a previous amplification reaction herein using the same or different primers may be so utilized. The specific nucleic acid sequence to be amplified may be only a fraction of a larger molecule or can be present initially as a discrete molecule, so that the specific sequence constitutes the entire nucleic acid.

It is not necessary that the sequence to be amplified be present initially in a pure form; it may be a minor fraction of a complex mixture, such as a portion of the β-globin gene contained in whole human DNA, or a portion of nucleic acid sequence due to a particular microorganism which organism might constitute only a very minor fraction of a particular biological sample. The starting nucleic acid may contain more than one desired specific nucleic acid sequence which may be the same or different. Therefore, the present process is useful not only for producing large amounts of one specific nucleic acid sequence, but also for amplifying simultaneously more than one different specific nucleic acid sequence located on the same or different nucleic acid molecules if more than one of the base pair variations in sequence is present.

The nucleic acid or acids may be obtained from any source, for example, from plasmids such as pBR322, from cloned DNA or RNA, or from natural DNA or RNA from any source, including bacteria, yeast, viruses, organelles, and higher organisms such as plants or animals. DNA or RNA may be extracted from blood, tissue material such as

chorionic villi or amniotic cells by a variety of techniques such as that described by Maniatis et al., Molecular Cloning (1982), 280-281.

The cells may be directly used without purification of the nucleic acid if they are suspended in hypotonic buffer and heated to about 90-100°C, until cell lysis and dispersion of intracellular components occur, generally about 1 to 15 minutes. After the heating step the amplification reagents may be added directly to the lysed cells. This direct cell detection method may be used on peripheral blood lymphocytes and amniocytes.

Any specific nucleic acid sequence can be amplified by the present process. It is only necessary that a sufficient number of bases at both ends of the sequence be known in sufficient detail so that two oligonucleotide primers can be prepared which will hybridize to different strands of the desired sequence and at relative positions along the sequence such that an extension product synthesized from one primer, when it is separated from its template (complement), can serve as a template for extension of the other primer into a nucleic acid of defined length. The greater the knowledge about the bases at both ends of the sequence, the greater can be the specificity of the primers for the target nucleic acid sequence, and thus the greater the efficiency of the process.

It will be understood that the word "primer" as used hereinafter may refer to more than one primer, particularly in the case where there is some ambiguity in the information regarding the terminal sequence(s) of the fragment to be amplified. For instance, in the case where a nucleic acid sequence is inferred from protein sequence information, a collection of primers containing sequences representing all possible codon variations based on degeneracy of the genetic code will be used for each strand. One primer from this collection will be homologous with the end of the desired sequence to be amplified.

The oligonucleotide primers may be prepared using any suitable method, such as, for example, the organic synthesis of a nucleic acid from nucleoside derivatives. This synthesis may be

performed in solution or on a solid support. One type of organic synthesis is the phosphotriester method, which has been utilized to prepare gene fragments or short genes. In the phosphotriester method, oligonucleotides are prepared that can then be joined together to form longer nucleic acids. For a description of this method, see Narang, S. A., et al., Meth. Enzymol., 68, 90 (1979) and U.S. Patent No. 4,356,270. The patent describes the synthesis and cloning of the somatostatin gene.

A second type of organic synthesis is the phosphodiester method, which has been utilized to prepare a tRNA gene. See Brown, E. L., et al., Meth. Enzymol., 68, 109 (1979) for a description of this method. As in the phosphotriester method, this phosphodiester method involves synthesis of oligonucleotides that are subsequently joined together to form the desired nucleic acid.

Automated embodiments of these methods may also be employed. In one such automated embodiment, diethylphosphoramidites are used as starting materials and may be synthesized as described by Beaucage et al., Tetrahedron Letters (1981), 22:1859-1862. One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Patent No. 4,458,066. It is also possible to use a primer which has been isolated from a biological source (such as a restriction endonuclease digest).

The specific nucleic acid sequence is produced by using the nucleic acid containing that sequence as a template. If the nucleic acid contains two strands, it is necessary to separate the strands of the nucleic acid before it can be used as the template, either as a separate step or simultaneously with the synthesis of the primer extension products. This strand separation can be accomplished by any suitable denaturing method including physical, chemical or enzymatic means. One physical method of separating the strands of the nucleic acid involves heating the nucleic acid until it is completely (>99%) denatured. Typical heat denaturation may involve temperatures ranging from about 80 to 105°C for times ranging from about 1 to 10 minutes. Strand separation may also be induced by an enzyme from the class of